Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 5389359, 15 pages
https://doi.org/10.1155/2022/5389359

Research Article

Privacy-Preserved In-Cabin Monitoring System for Autonomous Vehicles

Ashutosh Mishra, Jaekwang Cha, and Shiho Kim

School of Integrated Technology, YICT, Yonsei University, Seoul, Republic of Korea

Correspondence should be addressed to Shiho Kim; shiho@yonsei.ac.kr

Received 18 August 2021; Revised 28 October 2021; Accepted 21 March 2022; Published 22 April 2022

Academic Editor: Francisco Gomez-Donoso

Fully autonomous vehicles (FAVs) lack monitoring inside the cabin. Therefore, an in-cabin monitoring system (IMS) is required for surveilling people causing irregular or abnormal situations. However, monitoring in the public domain allows disclosure of an individual's face, which goes against privacy preservation. Furthermore, there is a contrary demand for privacy in the IMS of AVs. Therefore, an intelligent IMS must simultaneously satisfy the contrary requirements of personal privacy protection and person identification during abnormal situations. In this study, we proposed a privacy-preserved IMS, which can reidentify anonymized virtual individual faces in an abnormal situation. This IMS includes a step for extracting facial features, which is accomplished by the edge device (onboard unit) of the AV. This device anonymizes an individual's facial identity before transmitting the video frames to a data server. We created different abnormal scenarios in the vehicle cabin. Further, we reidentified the involved person by using the anonymized virtual face and the reserved feature vectors extracted from the suspected individual. Overall, the proposed approach preserves personal privacy while maintaining security in surveillance systems, such as for in-cabin monitoring of FAVs.

1. Introduction

Intelligent monitoring and surveillance systems are widely used to ensure safety and security. Popular applications of monitoring in public are video surveillance cameras (closed-circuit television); monitoring in intelligent transportation systems, including in-cabin monitoring and road traffic monitoring; and video monitoring for data generation and navigational tasks around city centers, airports, and public roads [1]. Driving automation also requires public visual information for multiple tasks [2]. The Society of Automotive Engineers defined six levels of autonomy in driving automation in 2014 (from no automation (level 0) to full automation (level 5)) [2–4]. Level 4 autonomous vehicles (AVs) are highly automated and capable of performing all driving tasks under certain conditions without human intervention. However, the driver (human) may control such AVs as and when required. In particular, fully autonomous vehicles (FAVs) (level 5 AVs) have no drivers; all occupants are passengers only [3, 4]. Therefore, no one oversees such AVs. In addition, in public and shared vehicles (such as ridesharing, carsharing, and car-full services in AVs), the passengers do not know each other. Therefore, it is important to ensure the security and safety of all occupants sitting in the cabin of such AVs. Furthermore, the vehicle should be protected from any malicious behavior of the occupants and/or external threats. Therefore, FAVs essentially require a multipronged in-cabin monitoring task in real time [5]. However, many countries have imposed a ban or severe restrictions on facial recognition techniques to secure personal information [6–16]. There are legal and ethical issues that impose various restrictions on public monitoring and surveillance systems [16–19]. Furthermore, identification of the accused is also important in abnormal (irregular) situations. This study was motivated by the fact that facial monitoring is important for safety; however, it poses a threat to individual privacy. In this study, we focused on the following two problems associated with in-cabin monitoring systems (IMSes):

(i) Protection of facial privacy.

(ii) Evidence of the accused in abnormal situations.

Therefore, a robust solution is required to provide privacy-preserved monitoring in public [20]. Moreover, it should be capable of identifying the concerned person when required. Figure 1 shows the dilemma of intelligent monitoring systems.

As illustrated in the above figure, an anonymous face protects personal information during in-cabin monitoring of an FAV. However, in certain irregular situations, personal identity is required to identify the accused person. An example of an abnormal incident or irregular situation can be an occupant of the FAV acting violently or attempting vandalism against the other occupants or toward the FAV itself. In such cases, it is important to identify the concerned person. Furthermore, this is an abnormal situation; however, in-cabin monitoring with real faces is not a solution to this problem. The breach of facial information leads to multiple consequences, such as misuse of facial data and banking and financial fraud [1, 6, 7, 13, 14]. One of our motivations for this work was to provide an approach that can protect against such problems in public monitoring systems, particularly the IMS. In-cabin monitoring with facial anonymization has security issues, while those with facial identity have privacy issues. Therefore, it creates a contradiction between privacy and security.

1.1. In-Cabin Monitoring. In-cabin monitoring is important in level 4 and beyond AVs [5]. It provides safety and security to the occupants. Simultaneously, it provides safety to the vehicle itself in an irregular situation. Past research works include in-cabin monitoring in various situations [21]. In-cabin monitoring for violence detection inside a FAV was reviewed in [22]. Bell et al. performed in-cabin monitoring to detect harsh vehicle maneuvers and risky driving behaviors [23]. Szawarski et al. patented the idea of in-cabin monitoring for a monitoring vehicle seat, occupants inside a vehicle, and the orientation of both the occupants and the vehicle seat [24]. Safety and cleaning problems of in-cabin monitoring of a vehicle were presented in [25]. However, a monitoring system should protect against any breach of personal privacy (facial identity) with the simultaneous ability to identify an actual person in case of irregular situations.

1.2. Facial Privacy versus Facial Recognition in Monitoring Applications. Real-time monitoring is essential in multiple monitoring applications. However, privacy in the public domain is an important concern in real-time monitoring tasks [26–30]. Facial anonymization is a common practice for preserving personal privacy. Recently, generative adversarial network- (GAN-) based deep learning (DL) models have been widely used for face swapping and anonymization [31–34]. In our previous study [31], we demonstrated a robust approach to preserving the facial identity of the occupants in a FAV cabin. It incorporated the facial swapping and reenactment technique to maintain privacy in in-cabin monitoring. However, in ab abnormal situation, the anonymized face of the occupants made it difficult to identify the concerned person [20].

1.3. Our Key Research Highlights. In this study, we propose an intelligent IMS. It is an efficient approach for identifying a person, even with an anonymized face. This method resolves both privacy and security issues. Accordingly, we can identify the person who causes an irregular situation, even with their anonymized face. In this approach, we preserved the key facial information of the occupants and stored these identity features on the cloud. These key features help in recognition of the person involved in the irregular situation. The highlights of this study are as follows:

(i) The concept of having an appropriate source face for each target face enhances puppeteering and reenactment of facial emotion and behavior. It helps in event and behavior detection in intelligent monitoring and surveillance systems in the public domain.

(ii) The involvement of the two-dimensional (2D) landmark position in the reenactment generator and separate segmentations of face and hair in the segmentation generator with inpainting and blending generators enhances the facial anonymization and reenactment operations.

(iii) The 128D identity feature is a key marker for accurate facial identification in an anonymized domain. The concept of storing a pair of IDs (original and anonymized) leads to reidentification without any privacy threat. It is not possible to know the original face with only 128D identity features. For reidentification, both the original visual input and the ID are required. In the cloud, the anonymized visual image with the original *ID* is stored. Therefore, there is no threat of privacy breach, even though the *IDs* are stored in the cloud.

(iv) Therefore, the proposed approach augments the facial identity feature information to locate the involved person in any abnormal situation without any personal privacy breach.

This approach pioneers a newer method of monitoring and surveillance to avoid any legal or ethical issues. Therefore, a monitoring database can be created in the anonymized domain, thereby facilitating further research on events and behavior monitoring in the public domain.

2. Materials and Methods

Personal privacy with identification is a challenge as well as a demand in real-time monitoring applications [20]. In this study, we developed a privacy-preserved IMS with the reidentification capability that can identify the accused person. The framework of the proposed method is shown in Figure 2. The proposed system operates in three stages. In stage 1, facial anonymization was performed to ensure personal privacy. It was performed using the onboard device of the

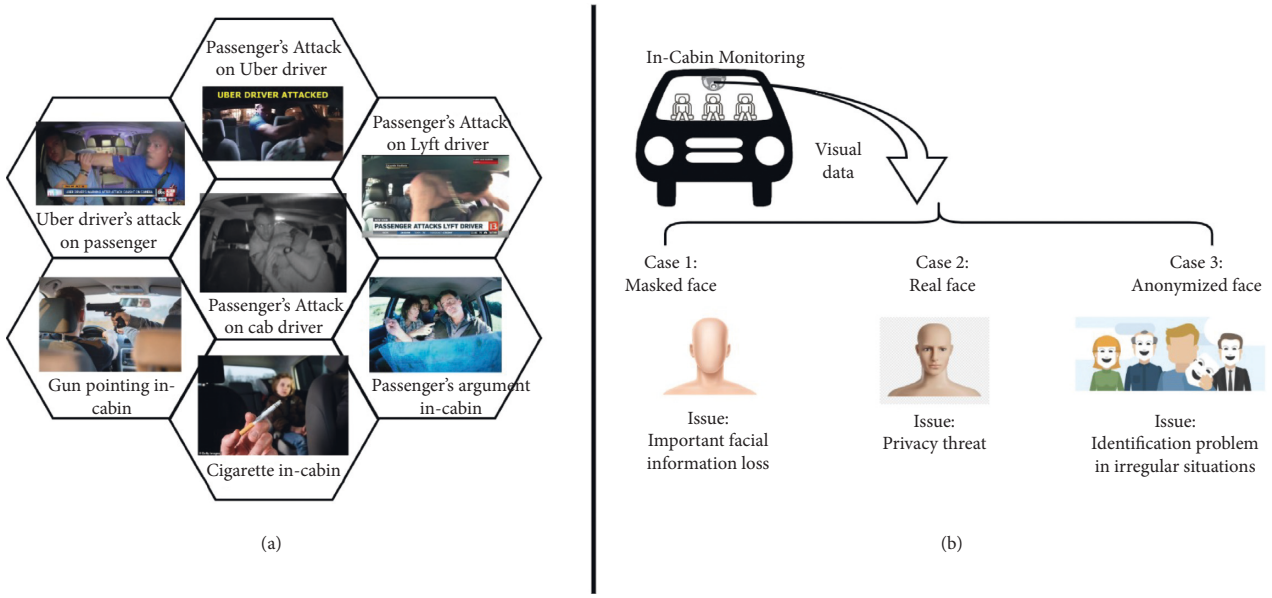

Figure 1: System overview of the proposed IMS. (a) Few examples causing abnormal situations in the cabin of a vehicle. (b) The dilemma of the legal and ethical issues (privacy) and practical problems (requirement of monitoring). Case 1: the masked face has no facial information, which is crucial in surveillance and monitoring inside the cabin of a vehicle. Case 2: real face suffers from personal privacy threats. Case 3: facial anonymization solves the problem of privacy; however, it has the problem of identifying the concerned person in case of irregular situations.

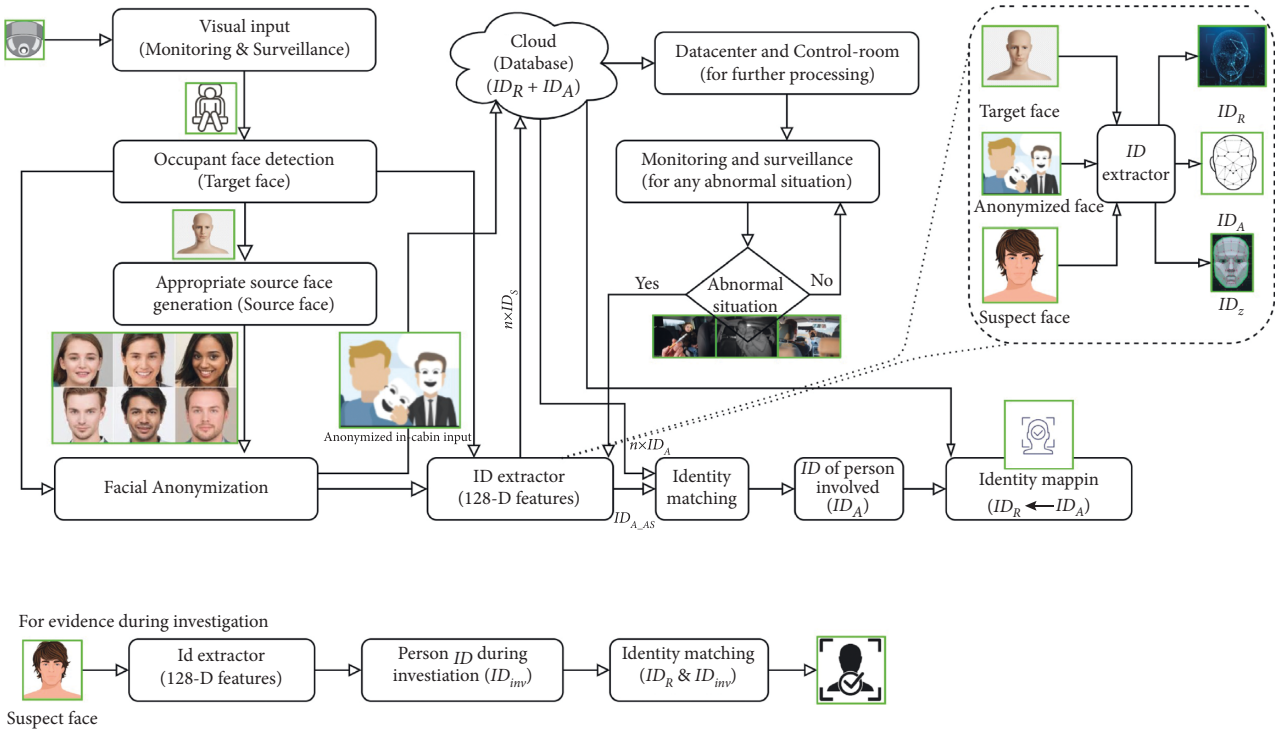

Figure 2: Proposed privacy-preserved intelligent IMS. Here, the identity features (*IDs*) are as follows: real face ID (ID_R), anonymized face ID (ID_A), ID of the occupant that caused an abnormal situation (ID_{A_AS}), and suspect face ID during the investigation (ID_{inv}).

AV. In stage 2, a pair of identity features (*IDs*) was generated for each face before and after anonymization (ID_R and ID_A). Further, the anonymized video along with the *IDs* was fed to the cloud. The pairs of *IDs* were kept in the cloud for person reidentification when required. The anonymized video frames were sent to the data center for further processing (monitoring and surveillance). In stage 3, the *IDs* were matched to search the accused (person involved in an

irregular situation (ID_{A_AS})). During the investigation, the similarity between *ID*s ensured the identification of the concerned person (*ID*). Further, during the investigation, this approach was verified by matching the *IDs* of the suspect face (ID_{inv}) at the time of investigation with the accused person's *ID*.

The dilemma between monitoring requirements and legal and ethical issues is also resolved through this approach. The details of the proposed approach are discussed thoroughly in Section 2.2. This approach is suitable for creating a monitoring and surveillance database with legitimation.

2.1. Materials. Many research works have been published on personal privacy and person identification considering these two issues as separate research problems. In this study, we briefly surveyed the related works and developments on both face anonymization and person identification.

2.1.1. Face Anonymization. Face deidentification preserves privacy-sensitive information. It alters the original face to hide privacy-sensitive information. Anonymization of faces is an easier and more robust solution to personal privacy-related threats in the digital domain [35]. Blurring, masking faces, or creating a patch over faces is slightly easier than any other face anonymization approach; however, those methods suffer from significant loss of facial information [32, 36]. Therefore, face swapping has attracted significant attention for facial anonymization purposes. The morphable model-based facial exchange approach is considered a pioneering work in face swapping [37]. Bitouk et al. demonstrated automatic face replacement in their work [38]. Machine-learning-based face swapping was suggested in [39]. A convolutional neural network (CNN) was used for face segmentation and swapping in [40]. GAN-based deep models have become popular for virtual human face generation [33, 34]. Therefore, along with autoencoders, GAN-based face swapping has gained considerable attention among researchers for seamless end-to-end face anonymization [33, 34, 41]. Face swapping-based automatic generation and editing of faces was showcased in [42]. It used a region-separative GAN (RSGAN). An autoencoder-based algorithm for face swapping was presented to detect fake videos [43]. In [44], a GAN-based encoder-decoder network was suggested to swap human faces. Collateral privacy issues have also been resolved using the face swapping method [45]. Nirkin et al. suggested a face swapping GAN (FSGAN) in [46]. It provided subject agnostic face swapping and reenactment between a pair of faces. Naruniec et al. presented a fully automatic neural face swapping method in [47]. Sun et al. proposed a hybrid model for face anonymization [36]. Hukkelas et al. introduced a GAN-based DeepPrivacy architecture for face deidentification to remove all privacy-sensitive information [34].

2.1.2. Person Identification. Facial recognition has multipurpose objectives, such as recognition, classification, and discrimination. Urbanization and smart cities demand widespread applications for face recognition [48–52]. Therefore, various face recognition approaches involving person identification have been demonstrated by past researchers. Face recognition approaches are classified into three categories: local, holistic, and hybrid approaches [52]. Local approaches involve only partial facial features (such as eyes, mouth, and nose) to recognize a face, whereas holistic approaches involve complete facial features, including background for facial recognition. Hybrid approaches, as the name suggests, involve both local and holistic approaches. In holistic approaches, popular algorithms involve independent component analysis, linear discriminative analysis, and principal component analysis [53, 54]. The development of artificial intelligence (AI) incorporating DL and CNNs has boosted the performance of facial recognition algorithms. Taigman et al. presented a deep neural network-based face recognition system, *DeepFace* [55]. Furthermore, many other extended versions of *DeepFace* have been demonstrated in multiple studies [56–59]. Adjabi et al. thoroughly reviewed face recognition techniques and their comparisons and future scope in their study [51]. Kortli et al. surveyed popular face recognition techniques in all three categories, that is, local, holistic, and hybrid approaches, in their study [52]. They compared these techniques in terms of accuracy, complexity, and robustness. They also discussed the advantages and disadvantages of the respective approaches. Wang et al. efficiently surveyed DL-based face recognition techniques in their study [60]. They exhaustively reviewed various popular DL-based approaches, including autoencoder-based, CNN-based, and GAN-based techniques. They also enumerated the key features, advantages, and disadvantages of these techniques. Furthermore, they summarized some of the commonly used datasets for deep face recognition. Moreover, they indexed the emerging real-world issues and major technical key challenges in deep facial recognition.

However, an application involving person identification must address important privacy concerns [61]. In particular, facial identification in the public domain must tackle individual freedom and ethics-related issues [51, 62]. Therefore, the state-of-the-art research problem in face recognition is the reidentification of an individual on anonymized data. Rocher et al. demonstrated the likelihood of correctly reidentifying a specific individual, even with the anonymized dataset [30]. They suggested a generative graphical model that can be trained on incomplete data to accurately identify individuals. Rooijen et al. suggested 2D video tracking for the reidentification of individuals in an anonymized dataset [20]. They suggested that the real facial information of a person is not necessary for reidentification. Luo et al. suggested effective training tricks for person reidentification [63]. A residual learning framework using the residual network (*ResNet*) model was suggested in [64] for visual recognition tasks. This facilitated the easier and more efficient training of a substantially deeper network. Schroff et al. suggested unified embedding using only 128 bytes per face for efficient face recognition [65]. They developed their network by incorporating the batch input

layer and deep CNN, followed by normalization. They used triplet loss to minimize the training errors. The world's simplest face recognition library (Dlib face recognition) is a popular and efficient tool for extracting facial landmarks [66]. It is a cross-platform open-source machine-learning toolkit that supports the development of machine-learning algorithms. It helps in recognizing and manipulating faces. Intent and behavior have been successfully detected using various techniques. Facial gesture sensing is performed using virtual reality (VR) and augmented reality (AR) devices, respectively in [67, 68]. AR/VR devices provide sensor responses to detect the intent or behavior of the user. However, FAV in-cabin monitoring requires intent or behavior detection using visual (computer vision (CV)-based) monitoring approaches.

2.2. Method. In this study, we proposed a representation learning-based approach to generate the identity signature of occupants. This signature is capable of deidentifying a person concerned with an irregular situation in the cabin of level 4 and beyond AVs. We proposed facial anonymization and reidentification system to provide countermeasures in case of an irregular situation. Therefore, this method provides personal information security with traces of the concerned person in case of any abnormality. The proposed method includes four main tasks. First, face anonymization with reenrollment. This is performed by using the face agnostic face swapping technique. It uses a set of GANs. These GANs are used for three purposes: facial reenactment and segmentation, facial inpainting, and facial blending. After accomplishing face anonymization, the second task is to extract the facial identity features of the occupant's faces in pairs (before and after anonymization, i.e., ID_R and ID_A) using the *ResNet*-based model. These *IDs* are stored in the cloud, and the anonymized video frames of in-cabin monitoring are transferred to the data center via the cloud for further processing. The third task of the proposed approach is to identify the accused by identity feature matching. Similarity matching of the *ID* of the accused obtained at the data center with the *IDs* of the occupants stored in the cloud ensures the identification of the concerned person (ID_A). However, it is the *ID* of the anonymized face of the accused. The Euclidean distance metric was used for similarity matching. Similarly, using the stored pairs of *IDs* (IDR and ID_A), we can obtain the real face identity feature of the accused (ID_R). Finally, in the fourth task, the evidence of the accused is obtained by matching the similarities between the *IDs* of the suspects with the *ID* of the accused during an investigation. Further details of the proposed method are provided in the following sections.

2.2.1. Facial Identity Feature Vector. The facial identity feature is (128, 1)-dimensional encoding of a facial image. It contains the encoded landmarks of the face using the *ResNet* model. The FaceNet-based CNN model and Facedlib face recognition library are used to extract the 128D identity features (*ID*) from the faces. Additionally, 128D is optimal embedding, which results in appropriate features required for reidentification or measuring the similarity between two faces. It has already been validated in the "*FaceNet*" architecture that fewer than 128D identity features deteriorate the identification performance; however, increasing the dimension only unnecessarily increases the number of parameters. This is the main reason for adopting the 128D identity features for recognizing faces.

Figure 3 shows the (128, 1)-dimensional facial identity feature vector generation of the occupant's face image. It uses a *ResNet*-based architecture consisting of 29 convolutional layers for this purpose. The *ResNet* architecture facilitates the dipper layer accessibility. Additionally, they have an inherent tendency to minimize the training error loss by increasing the number of layers. The triplet loss function is used to estimate the error in the reidentification of the concerned person. It performs similarity matching on the 128D identity features. For the anonymized anchor image *ID* (I_A), positive anonymized image *ID* (I_P), and negative anonymized image *ID* (I_N), the triplet loss is estimated by the following equation:

$$\mathscr{L}(A, P, N) = \max\left(\|I_A, I_{P2}\| - \|I_A, I_{N\|2} + \text{margin}, 0\right). \quad (1)$$

The anonymized anchor image ID (IA) represents the 128D ID of the person figured out in an irregular situation. The positive anonymized image (IA) is the stored image 128D ID of the same person on the cloud, and the negative anonymized image ID (IA) is the 128D ID of another occupant. Here, $(\|x, y\|_2)$ denotes the "Euclidean distance" between pairs $\{x, y\}$ in the triplet loss function. A factor margin is included in equation (1) to reduce the chances of misclassification. These facial features are incorporated in 128D encoding and are used as the facial recognizer using only 128 bytes per face.

Furthermore, a distance-based classifier compares the 128D features to identify the person involved in an irregular situation. It represents the difference between two feature vectors in Euclidean space. Suppose that image (R) represents the person. Image (C) is the stored image (copy) of the same person on the cloud, and image (D) is an image of another occupant. Further, $f(x)$ represents the 128D encoding of the image $f(x)$. The similarity (S) in the vector space is measured by the following equation:

$$S = \min\left(\|f(R), f(C)\|_2, \left(\|f(R), f(D)\|_2\right)\right). \quad (2)$$

It guarantees that images (R) and (C) are of the same occupant and are different from image (D), which is the image of another occupant.

2.2.2. Source Image Generation. A source image was required for face swapping in facial anonymization. It is used to replace the face appearing in the target image. This replacement, that is, swapping, should produce a realistic result that seamlessly reenacts the anonymized face that is similar to the target face. Our recommendation is to use a nonreal face as the source image. It mitigates any chaos/conflicts that may occur by using any real face as the source image. Therefore, in our proposed method, we used GAN-

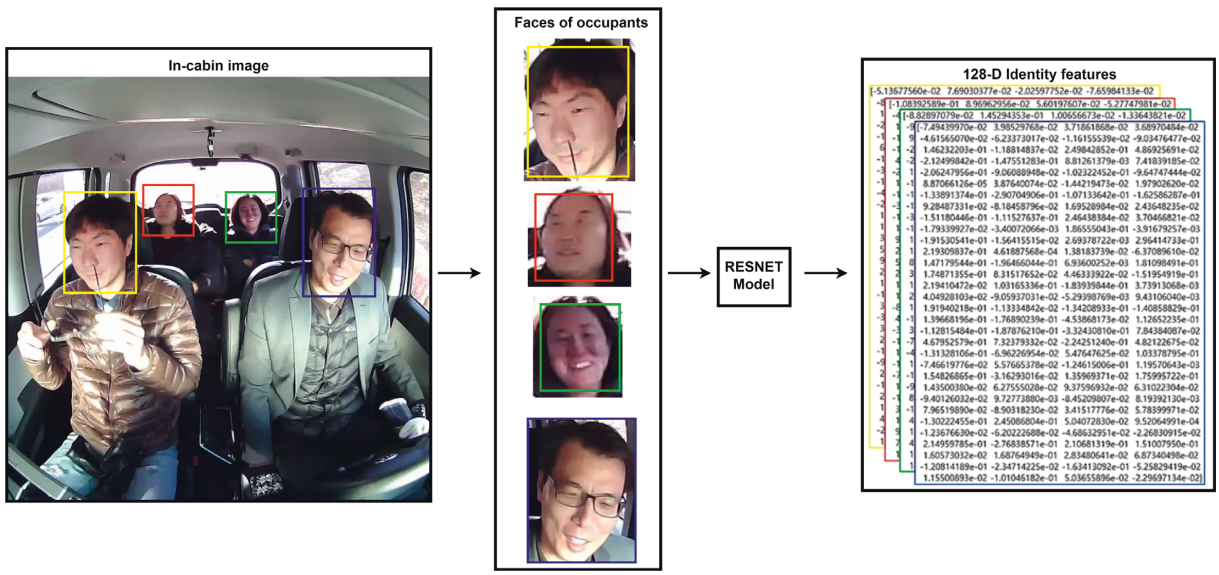

FIGURE 3: Illustration of 128D facial identity feature vector generation (from the occupant's face image). Image shown is taken from our in-cabin monitoring database. The numerical values in the yellow, red, green, and blue colored boxes are representing respective passengers' (128, 1)-dimensional facial identity feature vectors (*ID*).

generated virtual human faces as the source image. We have considered generating appropriate source faces that can effectively render the original emotions or behaviors performed by the occupants. It helps in further event and behavior-monitoring tasks. Figure 4 shows the proposed source image generation process. We applied the concept of similarity matching in vector space to select a similar source face for each target face from the set of virtual human faces (nonreal face as the source image). Similarity matching between source and target faces facilitates reciprocating similar emotions and intents, which is necessary for further monitoring applications.

Figure 4 shows the source image generation process. The face detector detects the faces (target faces) of the occupants (from the in-cabin visual input). The identity feature extractor extracts the *IDs* (128D identity features) of faces (target faces) and matches the similarity of the target faces with the set of virtual human faces (source faces) to find the most appropriate source face. This similarity matching is in the vector space (Euclidean distance matching between the extracted face *ID* and *IDs* of the set of virtual human faces).

2.2.3. Facial Anonymization. Facial anonymization requires exactitude in the anonymized faces to mitigate errors in further processing. Therefore, swapping should be performed efficiently to provide unaltered expressions and emotions over the anonymized face. We used the concept of FSGAN for facial anonymization to provide personal privacy during in-cabin monitoring of irregular situations. This requires perfection in the following three tasks:

(i) Facial Reenactment and Segmentation. To obtain proper facial swapping, we must estimate the proper reenacted face. This is performed by the proper segmentation of the face and hair segments of the target image. Proper facial reenactment requires separate face and hair segmentations with the mapping of 2D facial landmark positions. Therefore, the stepwise loss function is considered as the objective function for implementing facial reenactment. For i^{th} layer feature map ($F_i \in \mathbb{R}^{C_i \times H_i \times W_i}$), the perceptual loss ($\mathscr{L}_{\text{perc}}$) between pairs of images (x, y) is expressed as follows:

$$\mathscr{L}_{\text{perc}}(x, y) = \sum \frac{1}{C_i \times H_i \times W_i} \times \|F_i(x), F_i(y)\|_2. \tag{3}$$

The reconstruction loss ($\mathscr{L}_{\text{rec}}$) between a pair of images (x, y) is expressed as follows:

$$\mathscr{L}_{\text{rec}}(x, y) = \lambda_{\text{perc}} \times \mathscr{L}_{\text{perc}}(x, y) + \lambda_{\text{pixel}} \times \mathscr{L}_{\text{pixel}}(x, y), \tag{4}$$

where "λ" is the corresponding hyperparameter ($\lambda_{\text{perc}} = 1$; $\lambda_{\text{pixel}} = 0.1$; $\lambda_{\text{adv}} = 0.001$; $\lambda_{SG} = 0.1$; $\lambda_{\text{rec}} = 1$; $\lambda_{\text{stepwise}} = 1$) and $\lambda_{\text{reenactment}}$ is linearly increased from 0 to 1 during training. Pixelwise loss ($\mathscr{L}_{\text{pixel}}$) between a pair of images (x, y) is calculated as ($\mathscr{L}_{\text{pixel}}(x, y) = \|x - y\|$). We have used the multiscale discriminator adversarial loss objective function to improve the realism of the generated images. The adversarial loss ($\mathscr{L}_{\text{adv}}$) between the generator and discriminator (G, D) is expressed as follows:

$$\begin{aligned} \mathscr{L}_{\text{adv}}(G, D) &= \min\left(\max\left(\sum \mathscr{L}_{\text{GAN}}(G, D)\right)\right), \\ \mathscr{L}_{\text{GAN}}(G, D) &= E_{(x,y)}[\log D(x, y)] + E(x)[\log(1 - D(x, G(x)))], \end{aligned} \tag{5}$$

where "$E_{(x,y)}$" is the expected value over all real data instances. "$E_{(x)}$" is the expected value over all random inputs to the generator. The reenactment generator loss ($\mathscr{L}_{RG}$) is given by the following equation:

$$\mathscr{L}_{RG} = \mathscr{L}_{\text{perc}} + \mathscr{L}_{\text{rec}} + \mathscr{L}_{\text{adv}}. \tag{6}$$

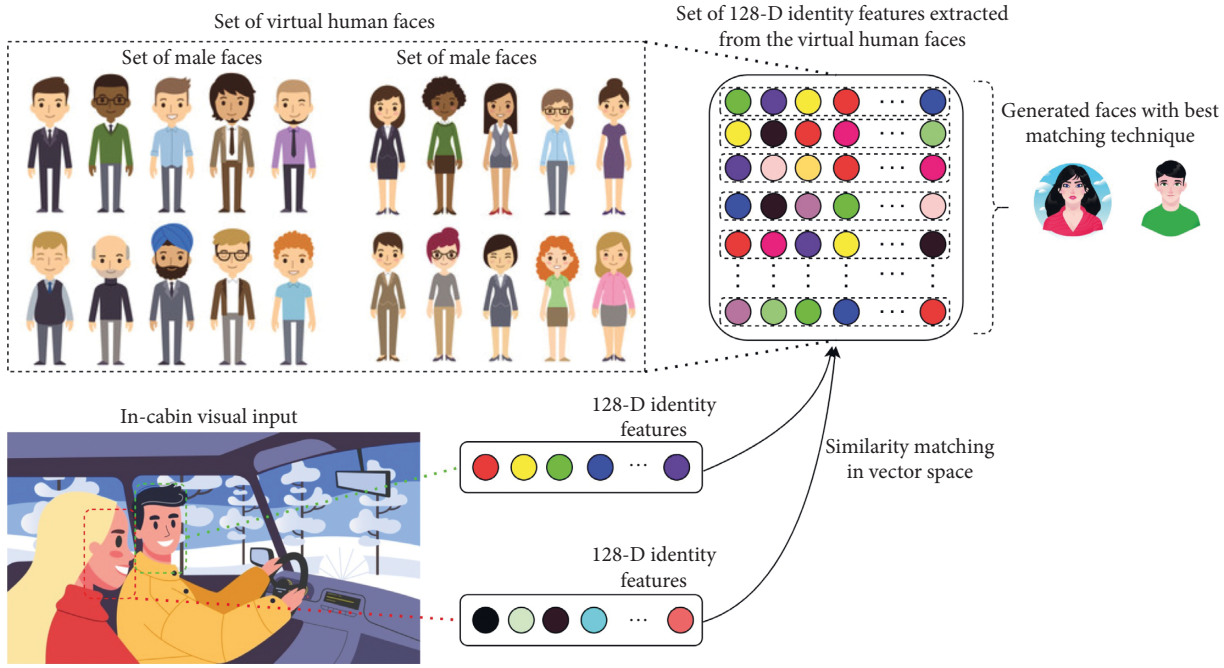

Figure 4: Source image generation using AI-generated faces with the best matching technique.

The perpetual loss is used to estimate the errors in capturing fine facial details, and the reconstruction loss is used to evaluate pixelwise color inaccuracy. Adversarial loss improves the generated images and provides a realistic look. The standard cross-entropy loss ($\mathscr{L}_{CE}$) is defined as (for truth label "t_i" and the "*SoftMax*" probability "P_i" for i^{th} class)

$$\mathscr{L}_{CE} = -\sum t_i \times \log(P_i). \tag{7}$$

Further, segmentation generator loss ($\mathscr{L}_{SG}$) is obtained by the following equation:

$$\mathscr{L}_{SG} = \mathscr{L}_{CE} + \mathscr{L}_{\text{pixel}}. \tag{8}$$

(ii) Facial Inpainting. This method estimates the missing portions of the reenacted face based on the face and hair segmentation of the target image. The inpainting generator loss ($\mathscr{L}_{IP}$) was calculated using the following equation:

$$\mathscr{L}_{IP} = \mathscr{L}_{\text{rec}} + \mathscr{L}_{\text{adv}}. \tag{9}$$

(iii) *Facial Blending.* It blends the completely reenacted face such that the swapped face matches the background environment like the original target face. The loss function ($\mathscr{L}_B$) for facial blending is obtained using the following equation:

$$\mathscr{L}_B = \mathscr{L}_{\text{perc}} + \mathscr{L}_{\text{adv}}. \tag{10}$$

The identity signature is generated corresponding to each occupant (a pair of identity signatures for real and anonymized faces) in the FAV. After facial anonymization, the video frames are transmitted to the cloud along with a pair of identity signatures of the occupants.

2.2.4. Anonymized Person Reidentification in Abnormal Situations. The proposed IMS facilitates the reidentification of the person involved in an abnormal situation. In our algorithm, in-cabin facial anonymization for preserving identity before transmitting the video frames to the cloud was achieved through the following pseudocode. The identity signature is generated corresponding to each occupant in the FAV. It is a vector of size 1×128. Therefore, for each occupant, we have a pair of identity signatures corresponding to the original and anonymized faces. Each pair is stored in the cloud. In any irregular situation, the concerned person is back-traced by matching the identity signature and anonymized face. The following is Pseudocode 1 of our proposed approach for obtaining the identity features (*ID*) of the person involved in an abnormal situation.

We considered virtual human face generation for the source faces. These faces are used to swap the target face in the captured visual in-cabin dataset. The source faces are generated depending on the similarity of the target face in the vector space. A similar source face provides the exactitude in replaying the facial gestures. This facilitates better reenactment performances. The concept of virtual human face generation for the source face protects any chaos or risk of threatening others' identities. Furthermore, we generated the facial identity signatures of the original and anonymized faces. These identity signatures help backtrack the concerned person in the event of an irregular situation. The identity signature is only vectored information. In other words, the identity signature in our proposed approach is extracted from a face that is used to reidentify the face. However, a face cannot be recreated using this information. Therefore, personal identity is not revealed through the identity signature. Our proposed approach provides proof or evidence that confirms the identity of the concerned person. The

(i) Definitions: Faces of the occupant (F); target face (T); appropriate source face (S); anonymized face (A); identity features (ID): real face ID (ID_R); anonymized face ID (ID_A); ID of the occupant that caused an abnormal situation (ID_{A_AS}); and an in-cabin abnormal situation (AS).
(ii) **Functions:** $\mathbb{F}$ = face detector; $\mathbb{S}$ = source detector; $\mathbb{A}$ = anonymizer; $\mathbb{I}$ = IDextractor.
(iii) Input: video frames (in-cabin)
(1) for $i = 1$ to range of the occupant:
(iv) $T(i) = \mathbb{F}$ (Input)
(v) $S(i) = \mathbb{S}(T(i))$ # search most similar source face for target face
(vi) $A(i) \leftarrow \mathbb{A}(T(i), S(i))$
(vii) $ID_R(i) = \text{II}(T(i))$ # 128D feature vector of the target face $ID_A(i) = \text{II}(A(i))$
(2) store: $ID(i) \leftarrow (ID_R(i); ID_A(i))$
(3) At datacenter: monitor event and behavior for AS:
(viii) if *occupant* j is involved in AS, then:
(xi) generate $(ID_{A\_AS}(j))$ #ID of j^{th} occupant in abnormal situation
(x) match ID:
(xi) for ID from 1 to range of the ID:
(xii) $k = argmin(\|ID_{A\_AS}(j), ID_A(:)\|_2)$
(4) Map: $ID_R(k) \leftarrow ID_A(k)$
(xiii) return $(ID_R(k))$ # the algorithm returns the real face ID of an anonymized person

Pseudocode 1: Algorithm for obtaining the ID of a person involved in an abnormal situation.

(i) Definitions: Target face (occupant's face) captured during the investigation (T_{inv}); ID of the occupant's face obtained during an investigation (ID_{inv}); ID of the person involved in the abnormal situation (ID_R); and real face of the occupant involved in the abnormal situation (O).
(ii) **Functions:** II = ID extractor.
(iii) Input: T_{inv}; ID_R
(1) At investigation:
(iv) for i from 1 to range of the target faces:
(v) $ID_{inv}(i) = \text{II}(T_{inv}(i))$
(vi) match ID: # compare ID_R and the suspect face ID
(vii) $j = \text{argmin}_i(\|\text{ID}_R, \text{ID}_{inv}(i)\|_2)$
(2) Map: $O \leftarrow j$
(viii) return (O)

Pseudocode 2: Algorithm for evidence of the person involved in the abnormal situation.

following is Pseudocode 2 of our proposed approach for evidence of the person involved in an abnormal situation.

In the case of proof or evidence, our method determines who is the concerned person. The returned identity feature (real face $ID_R(k)$) in Pseudocode 1 refers to the crucial identity parameter of the person involved in an abnormal situation. Matching the identity feature at the time of investigation with the obtained ID (real face $ID_R(k)$) confirms the person involved in an abnormal situation. Therefore, this approach easily locates the person involved in an irregular situation without any breach of others' identities.

3. Results and Discussion

In our experiment, we first anonymized the occupants of the FAV to secure their privacy in the public domain. Further, we applied the concept of vector space similarity to match the representation learning-based identity features for face recognition to locate the person involved in an irregular situation. The augmentation of the representation-learning-based identity feature introduces a new domain in re-identification. The proposed system was introduced to maintain personal privacy during the monitoring. We examined our proposed system for the in-cabin monitoring task of the FAV. We captured our database for in-cabin monitoring in abnormal situations. The similarity measure ($S_{i,j}$) is calculated by the Euclidean distance (ED) metric that is expressed as follows:

$$S_{i,j} = \|f(i), f(j)\|_2, \tag{11}$$

where $f(i)$ and $f(j)$ represent the 128D encoding of images i and j, respectively. Therefore, the similarity measure identifies the distance (Euclidean distance) between two pairs of *IDs* (128D encoding). The lesser the distance is, the closer the faces are.

3.1. Appropriate Source Faces. We proposed the concept of an appropriate source face in our facial anonymization approach. For every occupant face (target face), an

appropriate source face is obtained by matching their similarity in the vector space. We considered various scenarios to assess the efficacy of our proposed approach, including single and multiple faces in the input image frame. Figure 5 shows the complete set of the considered source faces in our experiment. We considered a set of 24 source faces (shown below). All these faces were not real (AI-generated). The source faces were used to swap the target face in the facial anonymization process.

These faces are nonreal virtual human faces. Generated Photos provides GAN-generated faces, which are human faces of nonreal humans. This has the benefit of further augmentation in anonymization. We considered various scenarios in our experiments. Examples include images with a single face only (for both males and females), multiple faces for males only and females only, and multiple faces for both males and females. These are in-cabin images obtained from the public domain (through an image search on the web) and are shown in Figure 6. We considered different scenarios for the occupants in the cabin. Therefore, in F1 and F2, there is only a single person in-cabin (F1: male and F2: female). In other scenarios, we considered more than one person in the cabin (only males, only females, and both males and females). Finally, we considered a family with children. There are four most appropriate source faces (S1 to S4) chosen for face anonymization.

Table 1 presents the similarities (in vector space) between the source and target faces, as shown in Figures 5 and 6.

These values follow the facial similarities of the source and target faces. These values measure the distance between the identity features of the source and target faces. The lower the values are, the more similar the faces are. The values in the green boxes represent the minimum Euclidean distances. These minimum values indicate appropriate source faces for anonymization. We can observe that the male target faces have lesser distances for male source faces than for female source faces. Interestingly, the distance values follow the similarity in looks as well. The eastern looks target faces have a lesser distance for eastern source faces than for the western source face, and vice versa. Female source faces have a lesser distance than the identity features of children's target faces.

3.2. Privacy Preservation during In-Cabin Monitoring. Facial anonymization is performed after deciding the appropriate source face using FSGAN-based face swapping and reenactment. Figure 7 depicts the reenacted anonymization of the target faces. Here, the first row (F1 to F8) and the third row (F9 to F23) show the original in-cabin visual inputs, and the corresponding anonymized output is represented in the second row (A1–A8) and fourth row (A9–A23). We chose four source faces (S1 to S4 shown in Figure 6) to swap the target faces (F1 to F23).

It is evident from this result that perfect reenactments are achieved even in the anonymized domain. Thus, it discerns the preservation of personal privacy during monitoring and surveillance operations. Furthermore, this appropriate reenactment supports the detection of abnormal or irregular situations in real time. To examine abnormality detection in the anonymized domain, we have experimented by considering vandalism as an irregular situation inside the vehicle cabin. We created our database for a similar situation. Snippets of the vandalism inside the vehicle are shown in Figure 8. We created a situation wherein occupants in the back seat of the vehicle started fighting with the occupants in the front seat. Four scenes were captured in our experiment. Shoulder shaking is shown in scene #1. Scene #2 shows a slapping scenario. Head shaking is discerned in scene #3, and scene #4 represents a neck choking incident inside the cabin of the vehicle. The identity features (*ID*s) of each occupant were calculated for normal and irregular situations. It is clearly observed that O3 (in the green box) is responsible for the irregular situation (in-cabin vandalism of the vehicle shown in the red box).

3.3. Person Reidentification in Abnormal Situations. Table 2 presents the similarities of the anonymized identity feature (ID_A) with the anonymized facial identity feature of occupant #3 (ID_{A_IS}). Here, ID_{A_IS} is the anonymized identity feature of the occupant who is involved in an irregular situation calculated at the data center, and ID_A is the anonymized identity feature of the occupant stored in the cloud.

The values in the green boxes represent the minimum Euclidean distances. These minimum differences between the *ID*s indicate the involved person. The original *ID* of this person is stored in the cloud. Therefore, by mapping the *ID*, we can easily identify the real person. Reidentification was performed by backtracking the *ID* obtained from the cloud and pictures of the occupants taken during the investigation. The *ID* of the person involved in an abnormal situation from the cloud (ID_R) needs to be matched with the *ID*s of the occupants inside the vehicle for facial identification of the person. This approach provides proof or evidence confirming the identity of the concerned person. For assurance of the person involved in the abnormal situation, we took pictures of the occupants (during an investigation). The images are shown in Figure 9. Now, the identity feature of each occupant is extracted to match the concerned person *ID* (ID_R) (as per Pseudocode 2). First, we compared the similarity between the faces of the occupants inside the vehicle with those of the other faces captured during the investigation. This is required to ensure that the occupants are the same.

Table 3 presents the similarity measures between the occupants' *ID*s extracted during an investigation and their *ID*s extracted from the in-cabin images.

The minimum Euclidean distances are represented by the green boxes. Here, minima indicate that the occupants O and O′′ are the same. Thereafter, assurance of the involved person is performed by matching the identity feature of the occupants extracted from the in-cabin image of the vehicle with the *ID* of the person involved in an abnormal situation (stored in cloud ID_R). Table 4 presents the similarity measures between the occupants' *ID*s extracted from the in-cabin image with the obtained *ID* of the person involved in an abnormal situation (stored in cloud ID_R).

Figure 5: Set of virtual human faces (AI-generated faces). These virtual human faces are obtained from Generated Photos. It provides AI-generated images that are free from any copyrights, distribution rights, and infringement claims (source: Generated Photos (https://generated.photos/)).

Figure 6: We have chosen single and multiple faces in the input images in different scenarios: single face (only male or only female), multiple faces (only male), multiple faces (only female), and multiple faces (both male and female). Here, the target (occupant) faces are indexed from F1 to F23, and considered source faces (both male and female) are indexed from S1 to S4.

Table 1: Similarities between the source and target faces.

Scenario		Target (occupants†)	Similarity measure (using Euclidean distance)			
			S1	S2	S3	S4
Single face	Male	F1	0.91489481	0.80287961	0.89433056	**0.74069120**
	Female	F2	0.78818484	0.81636149	**0.68592050**	0.76422129
	Male	F3	0.91414391	0.88400388	0.87615788	**0.83486502**
		F4	0.79685733	**0.71862379**	0.93450242	0.75311709
	Male	F5	0.91205174	0.82094296	0.87266242	**0.78036428**
		F6	0.81236709	0.80381698	0.93859941	**0.67143296**
	Female	F7	0.81097788	0.82709409	**0.71891988**	0.86480495
		F8	0.85947196	0.78512872	**0.77978978**	0.90500906
Multiple face	Both	F9	0.89428390	0.83158545	0.88401185	**0.80949051**
	Both	F10	0.84977716	0.90480697	**0.71174311**	0.94153902
	Both	F11	0.65831500	**0.52838455**	0.95610671	0.88142326
	Both	F12	**0.38916649**	0.45361382	0.89109294	0.88496564
	Both	F13	0.79624321	0.80097813	0.88202307	**0.71975099**
	Both	F14	**0.63660264**	0.67343593	0.88004248	0.96042187
	Both	F15	0.84524707	0.89008615	**0.77500429**	0.86828727
	Both	F16	0.74547080	0.77676084	0.93155677	**0.73583944**
	Both	F17	0.79179192	0.80390987	**0.73040828**	0.88839209
	Both	F18	0.78950908	0.79986798	**0.72049968**	0.94658813
	Both	F19	0.90007099	0.85322199	0.99829307	**0.85322199**
	Both	F20	**0.40197132**	0.63806865	0.83032199	0.88047087
	Both	F21	**0.46230089**	0.53098278	0.85879277	0.86241199
	Both	F22	**0.48055832**	0.54356384	0.83216505	0.81304802
	Both	F23	0.55751665	**0.45767183**	0.90473598	0.78517239

†The occupants are numbered from left to right clockwise.

Figure 7: Facial anonymization with reenactment. F1 to F23: original images. A1 to A23: corresponding anonymized images considering appropriate source faces.

Figure 8: Snippets of our database showing vandalism inside the vehicle cabin. The original image under normal and irregular situations is in row R1, and the corresponding anonymized images are shown in row R2. The occupants are numbered from left to right clockwise (O1, O2, O3, and O4). Scene #1: O3 shakes shoulder of O4; scene #2: O3 tries to slap O4; scene #3 O3 shakes head of O4; and scene #4: O3 chokes neck of O4. Green box: concerned person and red box: in-cabin vandalism.

Table 2: Identity feature matching between ID_{A_IS} #3 at the data center and other stored *ID*s of the occupants in the cloud for different scenarios.

Scene	Similarity measure (in Euclidean distance)			
	ID_A #1	ID_A #2	ID_A #3	ID_A #4
Scene #1	0.52893346	0.78363186	**0.35424358**	0.42124692
Scene #2	0.45234707	0.79687774	**0.35880417**	0.40963131
Scene #3	0.49863882	0.77615540	**0.41716736**	0.44500655
Scene #4	0.74701755	0.5816643	**0.53716927**	0.73605501

Detail description of scenes (scenes #1–#4) is mentioned in Section 3.2.

The zero value in the green box indicates that the occupant (O3) is the person involved in an abnormal situation. Overall, this approach focuses on in-cabin monitoring with personal privacy preservation to avoid abnormal situations. Personal privacy preservation is achieved by using the concept of event and behavior monitoring in an anonymized domain. The person's reidentification is only for providing evidence in cases where the involved person is denying it.

Figure 9: Other pictures of the occupants during investigation for matching. The numbering is the same as those in the in-cabin images from left to right (O1′′, O2′′, O3′′, and O4′′).

Table 3: Identity feature matching between the occupants' *ID*s extracted during the investigation and their *ID*s extracted from in-cabin images.

Occupant's *ID* (in-cabin)	Occupant's *ID*s were extracted during an investigation			
	ID_{O1}'	ID_{O2}'	ID_{O3}'	ID_{O4}'
ID_{O1}	**0.48432609**	0.79061829	0.72076523	0.69776956
ID_{O2}	0.75859866	**0.66392154**	0.79348079	0.72322158
ID_{O3}	0.52265982	0.77696899	**0.36226176**	0.47469558
ID_{O4}	0.64218059	0.81529880	0.52173335	**0.42871777**

Table 4: Identity feature matching between ID_R stored in the cloud with other occupant's *ID*s extracted from the in-cabin of the vehicle.

ID (person involved)	*IDs of the occupants (In-cabin)*			
	ID_{O1}	ID_{O2}	ID_{O3}	ID_{O4}
ID_R	0.62511349	0.75967812	**0**	0.52087737

4. Conclusions

Identity feature augmentation in anonymization is a potential solution for providing privacy in public domain monitoring. Identification of the involved person is crucial, especially in abnormal situations. The proposed intelligent IMS augments the security features with privacy. This method is suitable for creating a monitoring database without any restrictions or legalities. We performed various scenarios to assess the efficacy of the proposed system. It provided an efficient algorithm to perform monitoring tasks in the public domain without any threat to the personal identity of a person. This helped in reidentification, even with an anonymized face. In the future, this algorithm can be implemented on various public domain monitoring platforms, such as transportation systems, shopping centers, theaters, hospitals, highways, fuel refilling stations, smart city applications, and toll plazas.

Data Availability

The image data used to support the findings of this study are included in this paper.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Ashutosh Mishra and Shiho Kim contributed to the conceptualization. Ashutosh Mishra, Jaekwang Cha, and Shiho Kim developed the methodology, performed formal analysis, and investigated. Ashutosh Mishra reviewed and edited the paper. Ashutosh Mishra and Jaekwang Cha provided the software and performed validation, visualization, and data curation. Ashutosh Mishra provided the resources and preparation. Shiho Kim contributed to supervision, project administration, and funding acquisition. All authors have read and agreed to the published version of the paper.

Acknowledgments

This study was partially supported by the Brain Pool Program through the National Research Foundation of Korea

(NRF) funded by the Ministry of Science and ICT (NRF-2019H1D3A1A01071115) and by the Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (no. 2021-0-01352, Development of Technology for Validating the Autonomous Driving Services from the Perspective of Laws and Regulations). The authors thank all the volunteers for their valuable contribution to our database creation.

References

[1] P. Vennam, T. C. Pramod, B. M. Thippeswamy, Y.-G. Kim, and B. N. Pavan Kumar, "Attacks and preventive measures on video surveillance systems: a review," *Applied Sciences*, vol. 11, no. 12, Article ID 5571, 2021.

[2] J. Janai, F. Güney, A. Behl, and A. Geiger, "Computer vision for autonomous vehicles: problems, datasets and state of the art," *Foundations and Trends in Computer Graphics and Vision*, vol. 12, no. 1–3, pp. 1–308, 2020.

[3] "SAE international releases updated visual chart for its "levels of driving automation" standard for self-driving vehicles," February 2021, https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-%E2%80%9Clevels-of-driving-automation%E2%80%9D-standard-for-self-driving-vehicles.

[4] "Automated vehicles for safety," 25 February 2021, https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety.

[5] A. Mishra, J. Kim, D. Kim, J. Cha, and S. Kim, "An intelligent in-cabin monitoring system in fully autonomous vehicles," in *Proceedings of the 2020 International SoC Design Conference (ISOCC)*, pp. 61-62, Yeosu, Korea, October 2020.

[6] "UK's facial recognition technology 'breaches privacy rights," February 2021, https://www.theguardian.com/technology/2020/jun/23/uks-facial-recognition-technology-breaches-privacy-rights.

[7] "Facial recognition technology privacy and accuracy issues related to commercial uses," February 2021, https://www.gao.gov/assets/710/708045.pdf.

[8] "Facial recognition technology fundamental rights considerations in the context of law enforcement," February 2021, https://fra.europa.eu/en/publication/2019/facial-recognition-technology-fundamental-rights-considerations-context-law.

[9] P. Climent-Pérez and F. Florez-Revuelta, "Protection of visual privacy in videos acquired with RGB cameras for active and assisted living applications," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 23649–23664, 2021.

[10] F. Bignami, *Schrems II: The Right to Privacy and the New Illiberalism*, Verfassungsblog, Germany, 2020.

[11] D. Dushi, *The Use of Facial Recognition Technology in EU Law Enforcement: Fundamental Rights Implications*, Global Campus Open Knowledge Repository, Oxford, UK, 2020.

[12] A. Mekrani, *The Future of Facial Recognition in Relation to Privacy," Master Thesis*, Tilburg University, Tilburg, Netherlands, 2020.

[13] D. Naranjo, *Your Face Rings a bell: How Facial Recognition Poses a Threat for Human Rights*, Global Campus Open Knowledge Repository, Oxford, UK, 2020.

[14] "How facial recognition technology threatens basic privacy rights," February 2021, https://www.computerweekly.com/feature/How-facial-recognition-technology-threatens-basic-privacy-rights.

[15] M. Doktor, "Facial recognition and the fourth amendment in the wake of carpenter v. United States," *University of Cincinnati Law Review*, vol. 89, no. 2, p. 552, 2021.

[16] A. Daly, "Privacy in automation: an appraisal of the emerging Australian approach," *Computer Law & Security Review*, vol. 33, no. 6, pp. 836–846, 2017.

[17] M. Smith and S. Miller, "The ethical application of biometric facial recognition technology," *AI & Society*, vol. 37, no. 1, pp. 1–9, 2021.

[18] R. Van Noorden, "The ethical questions that haunt facial-recognition research," *Nature*, vol. 587, no. 7834, pp. 354–358, 2020.

[19] "Facial-recognition research needs an ethical reckoning," July 2021, https://www.nature.com/articles/d41586-020-03256-7.

[20] A. V. Rooijen, H. Bouma, R. Pruim, J. Baan, W. Uijens, and J. V. Mil, "Anonymized person re-identification in surveillance cameras," in *Proceedings of the Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies IV*, p. 115420A, International Society for Optics and Photonics, Edinburgh, UK, September 2020, http://toc.proceedings.com/56397webtoc.pdf.

[21] Y. Rong, C. Han, C. Hellert, A. Loyal, and E. Kasneci, "Artificial intelligence methods in in-cabin use cases: a survey," 2021, http://arxiv.org/abs/2101.02082.

[22] F. S. Marcondes, D. Durães, F. Gonçalves, J. Fonseca, J. Machado, and P. Novais, "In-vehicle violence detection in carpooling: a brief survey towards a general surveillance system," *Advances in Intelligent Systems and Computing*, vol. 1237, pp. 211–220, 2021.

[23] J. L. Bell, M. A. Taylor, G.-X. Chen, R. D. Kirk, and E. R. Leatherman, "Evaluation of an in-vehicle monitoring system (IVMS) to reduce risky driving behaviors in commercial drivers: comparison of in-cab warning lights and supervisory coaching with videos of driving behavior," *Journal of Safety Research*, vol. 60, pp. 125–136, 2017.

[24] H. Szawarski, J. Le, and M. K. Rao, "Monitoring a vehicle cabin," USPTO, Dallas, TX, USA, US. Patent 10252688, April 2019.

[25] X. Song, "Safety and clean vehicle monitoring system," USPTO, Dallas, TX, USA, US. Patent 10196070, February 2019.

[26] A. Taeihagh and H. S. M. Lim, "Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks," *Transport Reviews*, vol. 39, no. 1, pp. 103–128, 2019.

[27] D. J. Glancy, "Privacy in autonomous vehicles," *Santa Clara University School of Law*, vol. 52, no. 4, p. 1171, 2012.

[28] L. Collingwood, "Privacy implications and liability issues of autonomous vehicles," *Information and Communications Technology Law*, vol. 26, no. 1, pp. 32–45, 2017.

[29] H. S. M. Lim and A. Taeihagh, "Autonomous vehicles for smart and sustainable cities: an in-depth exploration of privacy and cybersecurity implications," *Energies*, vol. 11, no. 5, p. 1062, 2017.

[30] L. Rocher, J. M. Hendrickx, and Y. A. de Montjoye, "Estimating the success of re-identifications in incomplete datasets using generative models," *Nature Communications*, vol. 10, no. 1, pp. 3069–9, 2019.

[31] A. Mishra, J. Cha, and S. Kim, "HCI based in-cabin monitoring system for irregular situations with occupants facial anonymization," in *Proceedings of the International*

Conference on Intelligent Human Computer Interaction, pp. 380–390, Springer, Daegu, Korea, October 2020.

[32] T. Nakamura, Y. Sakuma, and H. Nishi, "Face-image anonymization as an application of multidimensional data k-anonymizer," *International Journal of Networking and Computing*, vol. 11, no. 1, pp. 102–119, 2021.

[33] S. Moschoglou, S. Ploumpis, M. A. Nicolaou, A. Papaioannou, and S. Zafeiriou, "3DFaceGAN: adversarial nets for 3D face representation, generation, and translation," *International Journal of Computer Vision*, vol. 128, no. 10-11, pp. 2534–2551, 2020.

[34] H. Hukkelås, R. Mester, and F. Lindseth, "DeepPrivacy: a generative adversarial network for face anonymization," in *Proceedings of the International Symposium on Visual Computing*, pp. 565–578, Springer, Lake Tahoe, NV, USA, September 2019.

[35] J. Dietlmeier, J. Antony, K. McGuinness, and N. E. O Connor, "How important are faces for person re-identification?," 2020, http://arxiv.org/abs/2010.06307.

[36] Q. Sun, A. Tewari, W. Xu, M. Fritz, C. Theobalt, and B. Schiele, "A hybrid model for identity obfuscation by face replacement," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 553–569, Munich, Germany, July 2018.

[37] V. Blanz, K. Scherbaum, T. Vetter, and H.-P. Seidel, "Exchanging faces in images," *Computer Graphics Forum*, vol. 23, no. 3, pp. 669–676, 2004.

[38] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 1–8, 2008.

[39] Y. Zhang, L. Zheng, and V. L. Thing, "Automated face swapping and its detection," in *Proceedings of the IEEE 2nd International Conference on Signal and Image Processing (ICSIP)*, pp. 15–19, Singapore, December 2017.

[40] Y. Nirkin, I. Masi, A. T. Tuan, T. Hassner, and G. Medioni, "On face segmentation, face swapping, and face perception," in *Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 98–105, Xi'an, China, 2018.

[41] T. Kim and J. Yang, "Selective feature anonymization for privacy-preserving image data publishing," *Electronics*, vol. 9, no. 5, p. 874, 2020.

[42] R. Natsume, T. Yatagawa, and S. Morishima, "RSGAN: face swapping and editing using face and hair representation in latent spaces," 2018, http://arxiv.org/abs/1804.03447.

[43] P. Korshunov, S. Marcel, and D. Fakes, "A new threat to face recognition? Assessment and detection," 2018, http://arxiv.org/abs/1812.08685.

[44] R. Natsume, T. Yatagawa, and S. Morishima, "FSNet: an identity-aware generative model for image-based face swapping," *Computer Vision - ACCV 2018*, vol. 11366, pp. 117–132, 2019.

[45] W. Bailer, "Face swapping for solving collateral privacy issues in multimedia analytics," in *Proceedings of the International Conference on Multimedia Modeling*, pp. 169–177, Thessaloniki, Greece, January 2019.

[46] Y. Nirkin, Y. Keller, and T. Hassner, "FSGAN: subject agnostic face swapping and re-enactment," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 7184–7193, IEEE, Seoul, Republic of Korea, August 2019.

[47] J. Naruniec, L. Helminger, C. Schroers, and R. M. Weber, "High -resolution neural face swapping for visual effects," *Computer Graphics Forum*, vol. 39, no. 4, pp. 173–184, 2020.

[48] A. K. Jain and S. Z. Li, *Handbook of Face Recognition*, Springer, New York, NY, USA, 2011.

[49] T. Huang, Z. Xiong, and Z. Zhang, *Face Recognition Applications, Handbook of Face Recognition*, pp. 371–390, Springer, New York, NY, USA, 2005.

[50] D. N. Parmar and B. B. Mehta, "Face recognition methods & applications," 2014, http://arxiv.org/abs/1403.0485.

[51] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, present, and future of face recognition: a review," *Electronics*, vol. 9, no. 8, p. 1188, 2020.

[52] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: a Survey," *Sensors*, vol. 20, no. 2, p. 342, 2020.

[53] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[54] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[55] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: closing the gap to human-level performance in face verification," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708, IEEE, Columbus, OH, USA, September 2014.

[56] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1891–1898, IEEE, Columbus, OH, USA, September 2014.

[57] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proceedings of the 27th International Conference on Neural Information Processing Systems (MIT)*, pp. 1988–1996, Montreal, Canada, June 2014.

[58] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2892–2900, Boston, MA, USA, June 2015.

[59] Y. Sun, D. Liang, X. Wang, and X. Tang, "DeepID3: face recognition with very deep neural networks," 2015, http://arxiv.org/abs/1502.00873v1.

[60] M. Wang and W. Deng, "Deep face recognition: a survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021.

[61] Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, I. Lagendijk, and T. Toft, "Privacy-preserving face recognition," in *Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium,(LNSC)*, pp. 235–253, Seattle, WA, USA, August 2009.

[62] A. Roussi, "Resisting the rise of facial recognition," *Nature*, vol. 587, no. 7834, pp. 350–353, 2020.

[63] H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, "Bag of tricks and a strong baseline for deep person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, March 2019.

[64] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[65] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: a unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern*

Recognition (CVPR), pp. 815–823, Boston, MA, USA, October 2015.

[66] D. E. King, "Dlib-ml: a machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[67] J. Kim, J. Cha, and S. Kim, "Hands-free user interface for VR headsets based on in situ facial gesture sensing," *Sensors*, vol. 20, no. 24, p. 7206, 2020.

[68] J. Cha, J. Kim, and S. Kim, "Hands-free user interface for AR/VR devices exploiting wearer's facial gestures using unsupervised deep learning," *Sensors*, vol. 19, no. 20, p. 4441, 2019.